DOCUMENT RESUME

ED 277 006 CS 210 183

AUTHOR Quellmalz, Edys S.; Burry, James
TITLE Analytic Scales for Assessing Students' Expository and Narrative Writing Skills. CSE Resource Paper No. 5.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE 83
NOTE 61p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Evaluation Criteria; Evaluators; *Expository Writing; Holistic Evaluation; Interrater Reliability; *Narration; *Rating Scales; Scoring; Secondary Education; Testing; *Writing Evaluation; Writing Research; *Writing Skills
IDENTIFIERS *Center for the Study of Evaluation CA

ABSTRACT

The Center for the Study of Evaluation's (CSE) expository and narrative rating scales have been developed to meet the need for instructionally relevant methods for assessing students' writing competence. Research indicates that large numbers of raters can be trained in the use of these scales and that, during training and independent rating, they can maintain high levels of agreement. The expository scale employs five elements to assess students' expository writing: general competence, essay organization and coherence, paragraph coherence, support, and grammar/mechanics. The narrative scale employs four elements to assess students' narrative writing: general competence, focus and organization, support, and grammar/mechanics. Each scale is (1) firmly anchored in research on writing and writing assessment, (2) intended for use in a setting in which students are provided with clear writing prompts containing explicit directions to help them to plan and develop their assignment, and (3) is numerically rated to assess specific criteria and examples of skill levels and deficient levels. Each scale also represents instructional priorities of the school system using the scales. CSE plans the training of raters carefully to ensure reliability in the scoring procedures. (The shorter, tabular version of each scale and copies of actual, scored student writing samples are appended.) (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED277006

ANALYTIC SCALES FOR ASSESSING STUDENTS' EXPOSITORY AND NARRATIVE WRITING SKILLS

Edys S. Quellmalz
James Burry

CSE Resource Paper No. 5

1983

CENTER FOR THE STUDY OF EVALUATION
UCLA Graduate School of Education
Los Angeles, California

CS 210 183

BEST COPY AVAILABLE

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Table of Contents

BACKGROUND AND PHILOSOPHY 1

OVERVIEW TO THE SCALES 3

The Theoretical and Research Base 3

The Expository Scale 8

The Narrative Scale 9

Features Common to Both Scales 10

The research basis 10

Setting the scope and focus of the assignment 10

The six-point rating scale 10

Adaptability to local curriculum needs 11

TRAINING AND PROCEDURES TO ENSURE RELIABILITY 14

Overview to the Problems and their Solutions 14

Training Sequence 15

Summary 21

CONCLUSION 22

REFERENCES 23

APPENDIX A: The CSE Rating Scales 24

APPENDIX B: Students' Writing Samples/Ratings 44

BACKGROUND AND PHILOSOPHY

The CSE expository and narrative rating scales have been developed to meet the need for sound, instructionally relevant methods for assessing students' writing competence. Each scale's analytic rating elements are referenced to specific rhetorical features characteristic of the writing aim addressed -- exposition or narration. This referencing is based on our belief that students' writing in each mode can be analyzed into its constituent elements. Knowledge of students' performance with respect to these elements can provide valuable information for assessing achievement and facilitating instructional planning at the classroom, school, and district levels.

The CSE analytic scales reflect our view of the roles of evaluation and testing in instruction. We are convinced that testing and instruction can become complementary activities and that both need to share the same educational goals and objectives. In the case of writing, for example, criteria used to assess students' compositions ought to reflect agreed-upon standards of good writing. Classroom instruction should address explicitly these same standards. In this kind of system, assessment can play a valuable role in instructional improvement by identifying specific areas of strength and weakness for individual students, and by indicating areas where classroom, school, and district curricula may need to be improved.

The CSE scales employ analytic rating procedures to meet the demands of the kind of assessment system described above. Raters assign points to each of several aspects of a student's composition, providing a rating of

the overall quality of the written product as well as ratings on specific elements. In contrast to the undifferentiated score provided by holistic scoring procedures, the CSE scales' overall rating and ratings on the specific elements give concrete information to guide instructional planning.

To be sure, some costs in rater time are involved in scoring written work analytically. Our studies indicate that it takes raters four to five minutes to assign analytic ratings to a multi-paragraph writing sample, and two to four minutes for a single paragraph. In comparison, it may take only about one minute to assign a holistic rating to a fairly short writing sample.

While we believe that the greater time commitment required by analytic rating is offset by instructional advantages, economics may require some compromises. For example, where student mastery and money are serious issues, the scale can be used to provide analytic ratings only for those students falling below mastery. In a similar vein, the scale can be abbreviated to target on specific skills of relevance to classroom instruction and ongoing classroom assessment.

In the remainder of this paper, we will provide an overview to the scales and describe the theory and research forming their base. We will describe the features, purposes, and uses of the expository and the narrative scales, and the training that CSE provides for school systems wishing to use the scales in their writing assessment programs.

OVERVIEW TO THE SCALES

We mentioned earlier that the CSE scales reflect the belief that students' writing can be analyzed into its constituent elements. We pointed out that if we can get accurate information about students' performance on these writing elements we can use it for instructional planning and improvement.

In the following sections we describe the theoretical and research basis for our belief that students' writing can be structurally analyzed. The work we describe led to the development of the analytic elements in our scales.

The Theoretical and Research Base

Writing is a very complex skill. While most people would probably agree with that statement, some people may assume that "a good writer is a good writer" and that a student who writes well in one mode will write equally well in another. The research points out some dangers in that assumption.

We can ask students to write in several modes of discourse or discourse aims. The two modes most commonly found in school curricula are expository and narrative writing. Expository writing involves expressing facts and ideas, and usually requires the student to support the ideas presented with appropriate detail, explanation, and logical development of thought. Narrative writing relates experiences and events, and often requires the student to provide chronological development and description of events.

We believe that a student called on to write in the expository mode will need to draw on different skills, reflecting different discourse structures, than he or she would need to achieve a narrative aim. For example, if a student is to explain something in the expository mode, he or she would need to present the main idea, develop the idea with supporting details, and perhaps summarize the issue presented. On the other hand, if a student is asked to give an account of something in the narrative mode, he or she might need to develop a sequence of events and describe the setting and characteristics with sufficient detail to make the account believable.

The research findings support our view of these differing skill needs. There is a good deal of evidence from studies of students' reading ability that different skills are needed to understand text written in different modes. There is similar evidence from studies of students' writing ability that different skills are needed to produce written text in different modes.

For instance, a student reading a passage written in the expository mode relies on a different set of organizing schemes, and different methods of breaking down, classifying, and understanding what is written than he or she would draw on to read a passage in the narrative mode (Meyer, 1975; Graesser, Hauft-Smith, Cohen, & Pyles, 1979).

Just as students use different processes to read and make sense of different kinds of writing, they also use different skills when they are asked to write in different modes such as exposition and narration (Veal & Tillman, 1971; Praeter & Padia, 1980; Quellmalz, Capell, & Chou, 1982.)

We mentioned earlier that we are interested in tying together testing and instruction. For other people sharing this interest the research offers a clear message: Different kinds of writing assignments require different kinds of student writing strategies and skills.

But there is a problem here. In many schools' writing programs, students are normally asked to compose in only one mode -- expository or narrative. Now, a student might have greater ability in writing a narrative assignment stressing chronological development than he or she has in writing an expository assignment requiring logical development. The opposite could also be true.

A teacher might get one kind of picture, therefore, about a student's writing ability based on how he or she develops narrative. The teacher might get a different picture about a student's writing ability based on how he or she develops an expository piece of work. If only one kind of writing ability is sampled, then it is possible that some writing deficiencies will go undiagnosed.

The point we are making here is simple: We cannot accept the assumption that a "good writer is a good writer." It depends on what we ask the student to write and for this reason we developed scales for the two discourse aims most frequently used -- expository and narrative (Quellmalz, 1980).

The CSE scales are important, therefore, because they provide information on students' expository and narrative abilities. They are equally important because of the way they assess these abilities.

There are two primary ways of assessing students' writing performance -- constructed response and selected response. A constructed response

provides a direct measure of a student's writing ability. Here, the student is asked to write something, perhaps a paragraph or two, perhaps an essay. When the piece of writing is turned in, a teacher or some other person judges the quality of the student's writing.

A selected response provides an indirect measure of a student's writing ability. Frequently, the student is given a passage to read followed by a multiple-choice test about the passage. This test may ask questions about the organization of the written passage, its supporting evidence, its grammar and mechanics. The score the student gets on the test is supposed to show how much the student knows about writing. And so far, maybe, so good. But there is a problem if we make the inference that student knowledge of writing quality accurately indicates student ability to produce good writing. If this inference is unjustified, then once again students' writing deficiencies may remain undiagnosed.

Several reasons are offered for using indirect, multiple-choice tests of students' writing ability. First, multiple-choice tests take less time to score than essays do. Second, some people believe that there is a strong relationship between students' multiple-choice test scores and their written work. Third, multiple-choice tests are more objective than ratings of students' written work. The argument is that if two people score the same essay in some general, impressionistic way, there is a good chance they will arrive at different judgments about the same piece of writing. Let us take up these three arguments.

Argument one: Scoring essays does take more time than scoring a multiple-choice test. However, our scales have been designed so that they do

not take up exorbitant amounts of rater scoring time. Depending on the length of the writing sample and rater familiarity with the scale, rating time ranges from two to five minutes. We believe the diagnostic and prescriptive information returned justifies the time invested.

Argument two: There is no guarantee that a student who scores well on a multiple-choice test of writing will be able to produce good writing. Researchers working specifically in writing do not believe that such indirect measures as multiple-choice tests provide an accurate indication of students' writing ability (Braddock, et al., 1963; Cooper & Odell, 1977; Quellmalz, Cappell, & Chou, 1982). If these tests are well constructed (and that is always a big if no matter what their purpose is) then they may be fair measures of reading comprehension, but not of how well a student will produce a piece of writing.

Argument three: The CSE scales have built-in procedures to make sure that different judges of a student's writing use the same decision rules. There is less liklihood, therefore, that they will give greatly different scores or grades to the same piece of work.

At CSE we built on some of the research mentioned above in our own work on the different writing and response modes (Quellmalz, Capell, & Chou, 1982). In our studies, we looked carefully at whether students' writing abilities in the two major modes -- exposition and narration -- were comparable. We also examined whether different response modes -- written work versus a multiple-choice test -- provided the same kind of information about student writing ability.

Here is a quick sketch of what we found:

° Students' writing skills did differ in the two discourse modes.

° While ratings of exposition were generally different than those of narration, the mechanics scores on the two modes were comparable. Perhaps this helps explain the false notion that "good writing is good writing."

° Multiple-choice scores were poorer indicators of student writing ability than measures based on actual student writing samples.

After these studies were conducted and the information analyzed, the scales we developed for the study were refined and now take the form we describe below.

The Expository Scale

This scale, developed as part of CSE's research on writing assessment (Quellmalz, 1980), is used to assess how well a student can write in the expository mode. By exposition, we have in mind writing that intends to inform by presenting facts, giving directions, recording events, interpreting facts, developing opinions. How well the student handles these elements influences the effectiveness of his or her expository composition.

The scale uses five elements to assess students' writing in the expository mode. These are: General Competence; Essay Organization and Coherence; Paragraph Coherence; Support; and Grammar/Mechanics. The first subscale, General Competence, is used for making a holistic evaluation of the essay's command of basic writing elements. The four other subscales focus on analytic evaluations of the quality of writing.

Using the General Competence subscale, raters read the composition to form a global judgement of how well the student arranged fundamental elements. In rating the composition for its general quality, raters keep

in mind the question of whether the work achieves an expository purpose for the intended audience. General Competence asks the rater to keep the question of fundamental arrangement in mind and give an overall rating showing how well the student handles the skills incorporated in the four remaining subscales. The rater can assign the composition a global score ranging from six to one. Scores of six, five, and four represent varying degrees of mastery; scores of three, two, and one represent varying degrees of non-mastery. This six-point system, which is described in greater detail later in the paper, applies to all the subscales.

In the remaining subscales, the rater's attention is directed to specific skills and qualities. Each skill or quality is described in detail.

- The Essay Organization and Coherence subscale focuses on the flow of ideas throughout the entire composition and between its paragraphs. Emphasis here is on vertical relationships throughout the essay. The rater is reading to see if the essay has a main idea, for example, which makes a point about the topic, whether that point is at a greater level of generality than the other points in the paper, and if the points made relate to the essay's thesis.

- The Paragraph Organization subscale is concerned with horizontal relationship within paragraphs, with the logical arrangement of points and their subordination to the paragraph topic.

- The Support subscale focuses on the specificity, depth, and amount of elaboration used to develop the theme.

- The Grammar/Mechanics subscale focuses on errors in sentence structure or mechanics and how seriously they interfere with communication. It pinpoints global errors making it difficult to understand the writer's message, and more local errors which are not serious impediments to communication.

The Narrative Scale

This scale, which was also developed as part of our research on writing (Quellmalz, 1980), is used for assessing how well a student writes in

the narrative mode. By narration, we have in mind a description of an event(s), based on an account of such aspects as (1) how, when, and where the event(s) took place; (2) the persons, places, or things involved in the event(s); and (3) the actions, thoughts, or feelings of the actors described. How well the student handles these features influences the effectiveness of his or her narrative essay.

The scale uses four elements to assess narrative writing. These are: General Competence; Focus and Organization; Support; and Grammar/ Mechanics. The first subscale, General Competence, is used to make an overall or holistic evaluation of the essay's command of writing fundamentals. The three other subscales are used to make analytic evaluations of the specific component features of narrative writing. Again, these three subscales focus on specific elements which are described in detail.

General Competence is based on a global judgement of the composition. The central question is the narrative effectiveness of the work and how well it arranges the features incorporated in the remaining subscales. A six-point scale, with the same values as those described for the expository scale, is used.

- The Focus/Organization subscale is concerned with the composition's structure and flow of ideas--within- and between- paragraphs.
- Support, which has a similar emphasis to that described for exposition, focuses the rater's attention on the development of events, descriptions, and characters through the use of well-integrated details.
- The Grammar/Mechanics subscale focuses on the kinds of global and local errors in sentence structure and mechanics described above.

Features Common to Both Scales

The research basis: Each scale, and its specific analytic elements, is firmly anchored in our own and others' research on writing and its assessment. The global or holistic judgment offered in each scale assumes that the whole is greater than the sum of its parts, and asks for a rating of the quality with which the writer engages the topic to achieve the intended purpose. The discrete, analytic features in each scale ask for separate ratings of a fundamental set of elements. These elements are specified because they recur in the research as basic features of acceptable writing. They also represent priorities in many writing competency programs. The individual subscales therefore present specific criteria for judging the student's skill in using these elements.

Setting the scope and focus of the assignment: Each scale is intended to be used in an assessment setting in which students are provided with clear writing prompts containing explicit directions to help them plan and develop their assignment. These directions should include: the specific purpose or function of the writing; a specific audience to whom the student will write; a specific topic or subject to write about; and criteria that will be used to judge the essay.

These directions are intended to provide students with a clear and common understanding of the task expected of them; that is, the rhetorical context. As such, they are the initial step in ensuring that each student's writing is judged on the same task description. They are a counterpart to the common scale criteria used by raters.

The six-point rating scale: Each numerical rating on the scale is tied to specific criteria and examples of skill levels and deficiency levels.

These numerical ratings are meaningful and consistent indicators of students' writing performance because they provide raters with specific guidance to decide upon the numerical rating that the composition should receive with respect to a given scale.

On a given subscale a student receives a score indicating his or her degree of mastery of the skills specified for that scale. In exposition, for example, for a student to receive a rating of six (highest mastery level) on overall organization and coherence, the rater is directed to look for certain features in the composition. Among these features are: limitation of the topic; logical essay plan; maintenance of essay plan.

For the student to receive a score of four (lowest level of mastery) on this part of the expository scale, among the features the rater is directed to look for are: clear main idea, logical plan, and reasonable support.

Each of the other subscales used for expository rating offers similar criterion levels and examples.

The same level of specificity of guidance to the rater is offered in the narrative scale. For example, on the support subscale, for a student's composition to receive a rating of five (second-highest mastery level) his or her work must provide supporting details, such as examples and descriptions, to develop events or characters. A student whose use of detail is not integrated would receive a score of two.

Adaptability to local curriculum needs: The subscales, with their accompanying operational criteria, should represent instructional priorities of the school system using the scales. In an ideally integrated assessment and instruction system, criteria used to evaluate writing

gathered for formal assessment should be the same as those used to evaluate and provide feedback to students on their more routine classroom writing assignments. While individual classrooms might focus on additional elements such as originality of content and style or voice, criteria used in judging formal, functional writing should reflect desired core writing elements. The CSE scales, therefore, include these commonly valued basic writing elements.

These core elements are intended to provide starting points for schools and teachers to consider as they plan their own writing assessment needs. For example, elaborated mechanics elements of sentence construction, usage, spelling, punctuation, and capitalization can provide useful classroom-level diagnostic information, especially for younger or less proficient students whose writing shows various mechanical weaknesses. On the other hand, for large-scale competency testing programs assessing older or more able students, a comprehensive mechanics rating may be sufficient.

As we will see in the next section, when CSE provides schools and districts with training in the use of the scales, part of that training involves making sure the criterion descriptions are agreed upon by the potential raters.

Copies of the complete, expanded scales and scoring criteria appear in Appendix A. This expanded version serves primarily as a "text" which raters read before training to familiarize themselves with the scales' rationale, structure, criteria, and criterion examples. During training and actual rating, raters use the shorter, tabular version of the scale, which appears at the end of Appendix A. Copies of actual student writing samples scored using our scales appear in Appendix B.

TRAINING AND PROCEDURES TO ENSURE RELIABILITY

In the preceding secion we mentioned that students need clear directions to make sure that they all bring the same understanding of the writing task to the job at hand. We also pointed out that raters need to follow the same criteria as they assess students' written work.

This section talks more about the procedures which help ensure that raters do indeed apply the same criteria. It also offers some reasons for why raters, even when they do try to use the same criteria, can "drift" apart as they go about the job of scoring students' written work. The training we offer helps overcome this problem.

Overview to the Problems and their Solutions

Samples of students' written work can provide direct evidence of writing ability. Scoring these samples, however, can present a problem, even with explicit criteria to follow; no matter how good the scoring system, assuring reliability or agreement among raters requires careful planning. First, raters who come to a training session with differing views of evaluation must accept and then learn how to apply the rating scales accurately and consistently. Second, raters who have achieved high levels of agreement during training tend to drift apart and show less agreement when they must score large numbers of papers. This drift may be caused by fatigue leading to careless application of criteria, or it may be the result of other influences, such as the range of quality in the papers being scored or the re-emergence of idiosyncratic rater values.

What this means, then, is that even when a careful rating system is used, the more papers a rater reads the more likely he or she will drift away from the intended criteria. The training we provide, which grew out of previous CSE research and technical assistance in writing assessment (Quellmalz, 1980), has built-in procedures to control this potential drift and to keep raters on track.

Training Sequence

The training sequence has three basic features. First, it emphasizes developing consensus in defining scale elements that are sensitive to local needs. For example, the expository scale may present separate subscales for focus and organization rather than a combined coherence scale. Also, a district can decide to add features of concern in more advanced writing, such as style or attention to audience.

Second, specific criteria are provided for each element, and raters receive repeated practice and feedback in their application. Third, the training serves as precursor to actual ratings of students' written tests; it is not merely an academic exercise. Therefore, the rating procedures are carefully structured to ensure that raters learn to apply criteria accurately and continue to do so as they go on to rate actual test essays independently. A key feature of this research-based training sequence is the use of essays which have been prescored by a panel of experts and which are then used as examples during training, as qualifying papers at the end of training, and as inserted "check" papers during independent scoring to identify drifting raters and to help them stay on track.

Training proceeds as follows:

Setting up a Rater Training Session

Step 1: CSE and the school or district asking for the training discuss overall purposes. Questions may include: Is training to be provided for one scale or both? Will the scales be used for routine classroom assignments, school/district competency testing, or both?

Step 2: The district collects representative student writing samples. These may come from pilot tests of prompts in neighboring districts. The samples must be available well in advance (three weeks) of the scheduled training session.

The samples should reflect the spectrum of student writing abilities and represent the grade levels tested and/or in which the raters teach.

The sample/papers are used for three purposes:

1. as practice training papers (approximately 30)

2. as pilot test papers to be sure that raters qualify at the end of training before proceeding to independent rating (approximately 20 to 30)

3. as check papers that will be inserted among the essays that will actually be judged during independent rating (a set of three papers to be read after every one or two hours of scoring)

Step 3: Using the CSE scale, CSE and district staff prescore the samples for use as training, qualifying, and check papers. "Solid" and marginal examples rated from 1 to 6 on each of the subscales are prepared. "Feedback sheets" presenting the scores and reason for the score that cite features of the papers and use language from the scale are prepared for all check papers and for about 10 of the training papers.

Step 4: Packets of the students' essays are prepared for each rater. The training packet contains copies of the expanded, prose version and the shorter, tabular version of the rating scales, and the training essays. To acquaint raters with the check procedure they will encounter later, written feedback sheets are prepared for about 10 of the training papers; trainers, however, have notes to explain all the prescored training papers. Feedback sheets will only be distributed after raters have scored a paper.

The qualifying packet consists of 10 to 20 prescored essays that represent the full range of student compositions.

The check papers may be put in separate packets for each rater or be available at a central location.

Test folders are prepared containing 10 essays which will be rated independently. The 10 papers in each folder are randomly selected from the full set of student essays to be rated. Each folder is numbered.

A rater assignment sheet is prepared listing the sequence of folders that each rater who has qualified will score. Each folder of essays will be scored by two raters. These raters are randomly assigned so that no two raters are systematically paired.

Sheets for recording the scores that raters give to training papers, qualifying papers, and actual test papers are prepared.

Providing the Rater Training

Step 1: The training begins with a brief description of the scale's structure and rationale. Discussion deals with the broad features of the scale to be used, its research basis, and the need to develop common understanding of purpose and to maintain high levels of rater agreement.

Copies of the expanded scale are passed out and read by the raters.

Step 2: The trainer and the raters discuss each element in the scale being used. Discussion includes the relationship between the scale's general competence rating and ratings on the subsequent subscales. Each of the subscales is defined and examples of the writing elements it focuses on are provided.

Discussion then focuses on the six-point rating system applied to each subscale. Particular emphasis is placed on the operational criteria provided for each possible rating. The point of this discussion is to ensure common understanding of each subscale and what each possible rating means. Any language clarification necessary is entered directly on the scale, which is then used for the remainder of the training. The trainer points out that levels of rater agreement which are lower than .80 (that is, where raters are in lower than 80% agreement) will suggest less than-uniform scale application.

Step 3: The trainer passes out the prepared training packages of student writing samples. Raters are directed to read and then rate the first three writing samples in their package. The major intention here is for raters to begin "practicing" the criteria on a few papers.

Step 4: Discussion begins after raters have scored the first three training papers. The trainer presents the scores that the experts gave to these papers and explains the reason for these scores by reading through the essay and noting the features that led to the score it received on each subscale.

Step 5: Raters rate another three papers, and the trainer displays the scores that raters gave each essay, using the following tabular format:

	Rating					
Subscale	1	2	3	4	5	6
_____						
_____						
_____						
_____						

The trainer then presents the expert scores for these papers, and asks one or two individuals who gave scores that are two or more points away from the expert score to explain their reasoning by identifying features in the papers that exemplify a score category described in the scale. The trainer then explains the basis for the expert scores.

Step 6: Depending on the level of agreement among raters after they have scored the first six papers, discussion may follow a variety of topics:

- is there need for further refinement of the language of the subscales?
- is everyone bringing a common framework to the use of the scales? do some raters have routine classroom diagnosis in mind while others are applying the scales from the standpoint of minimum competency?
- do these differing frameworks suggest more lenient/more stringent application of criteria? is that appropriate?

Step 7: Raters then continue to rate training papers in sets of five. After each set has been rated, the trainer distributes feedback sheets for two or three of the papers in the set. These papers are discussed only if questions are raised. Group discussion focuses on the other papers in the set, following the strategy described in Step 5, and emphasizes those subscales where agreement levels are less than .80, that is, where less than 80% of the raters are in agreement.

After 10 to 15 papers have been scored, discussion and focus may become more directive. For example, if only a few individuals are giving discrepant scores, the trainer may work with these individuals while they are scoring a set.

Step 8: When 80% agreement levels are reached on the third or so set of training papers, the trainer may decide that it is time to distribute the qualifying papers. This set contains 15 to 20 papers which raters score without interruption or discussion. To qualify, raters must agree with each other within one point on each subscale on at least 80% of the papers.

If most of the group qualifies and is ready to proceed, raters then begin the independent rating of the actual student test papers.

At the same time, additional training may continue on a subscale with less than 80% agreement. For example, individuals who do not agree at the 80% level may continue to practice and discuss papers individually or in small groups with the assistance of the trainer.

Independent Rating of Essays

Step 1: Raters refer to the rater assignment sheet to get the folders of essays they will score. After raters have scored one or two folders, they rate the set of three check papers and record the scores they gave them. They then turn in their check scores, and read the feedback sheets for the three papers. The point here is to determine if the pair of raters scoring the same set of test papers are (1) in agreement with each other and (2) in agreement with the scores given by the "expert" rater. Since students will actually be graded in the process, it is critical that raters' scores be no more than one point off the expert rater's score. Where greater differences exist, trainer and raters discuss the paper in question to resolve discrepancies.

Step 2: Depending on the number of actual student essays to be rated, independent rating may continue over a few days, perhaps for a few hours each day. In this case, at the beginning of each rating day, raters may begin by reading and rating a set of check papers. This serves to keep raters on track and to minimize the re-emergence of idiosyncratic criteria.

Summary

Routine assessment of student's classroom progress may not require the careful check process described above. But it is another matter when large numbers of raters are to judge the writing performance of large numbers of students, say in a minimum competency testing program. Here the check procedure is critical, given the possibility of rater disagreement emerging over time.

We strongly recommend, further, that when writing assessment results are to be used to make important decisions about student ability, each student's composition be read and scored, as described above, by two raters. If the two raters are in agreement at the .80 level or higher, the student's composition can be assigned the average of the two scores. If the two raters show agreement lower than .80, then the student's composition should be rated again by a third, independent judge. The two raters and the judge can then discuss differences via the kind of process described in Step 5 -- citing features of the subscale element in the student's composition that match or exemplify the scale criteria for the score given. This discussion should lead to resolution of the rating discrepancy.

CONCLUSION

In this paper we describe the need for using accurate and reliable criteria for the assessment of student writing. We offer scales for expository and narrative writing that we believe meet the standards of accurate and fair assessment.

Both our research in the assessment of writing and the technical assistance we have provided in several settings indicate that large numbers of raters can be trained in the use of the scales, and that during training and independent rating they can achieve and maintain high levels of agreement. Depending on the numbers of essays to be read and the numbers of people who will be working as raters, the kinds of check procedures we describe are critical to the achievement and maintenance of high levels of inter-rater reliability.

It may be, at least initially and until raters become familiar with the scale procedures, that significant investments of time will need to be made. However, in the long run, if we are concerned about linking assessment with instruction so as to provide information for instructional improvement, the time invested can lead to high returns.

For example, in schools where the CSE scales are used in the writing assessment program, teachers state that the training and the scales have helped them to plan and monitor instruction. They have become more aware of specific writing skills and are directing their instruction to scale elements that result in improved student writing.

REFERENCES

Braddock, R., Lloyd-Jones, R., & Schoer, I. Research in written composition. Urbana, Il.: National Council of Teachers of English, 1963.

Cooper, C., & Odell, L. (Eds.) Evaluating writing: Describing, measuring, judging. Buffalo: State University of New York at Buffalo, 1977.

Graesser, A.C., Hauft-Smith, K., Cohen, A.D., & Pyles, D. Familiarity and test genre on retention of prose. Journal of Experimental Education, 1979, 48, 281-290.

Meyer, B. F. The organization of prose and its effects on memory. North Holland Studies in Theoretical Poetics (Vol. I). Amsterdam: North Holland Publishing Company, 1975.

Praeter, D., & Padia, W. Effects of modes of discourse in writing performance in grades four and six. Paper presented at the annual meeting of the American Educational Research Association, Boston, 1980.

Veal, L.R., & Tillman, M. Mode of discourse variation in the evaluation of children's writing. Research in the Teaching of English, 1971, 5, 37-45.

Quellmalz, E.S. Controlling rater drift. Los Angeles: UCLA Center for the Study of Evaluation, 1980.

Quellmalz, E.S., Capell, F.J., & Chou, C.P. Defining writing: Effects of discourse and response mode. Journal of Educational Measurement, 1982, 14, 4.

APPENDIX A

The CSE Scales

Expository Scale

Global Rating Procedures

The purpose of global rating is to form a single judgement of how well a piece of writing communicates a whole message to the reader. Global scoring assumes that each characteristic that makes up an essay -- organization of ideas, content, mechanics, and so on -- is related to all other characteristics. It further assumes that some qualities of an essay cannot easily be separated from each other. In short, the procedure views a piece of writing as a total work, the whole of which is greater than the sum of its parts.

Discerning readers naturally will attend to, or be influenced by, some essay characteristics more than others. In this general scoring, however, readers should arrive at a judgment regarding the essay's overall quality.

For this element, you are being asked to form an overall judgement concerning the effectiveness of the essays as examples of expository writing. The judgement should consider all the elements on the scale -- essay organization, paragraph organization, support, and mechanics.

Some views on exposition are given below:

- Exposition is the kind of discourse that explains or clarifies a subject.
- Exposition seeks to explain or inform through such methods as giving reasons or examples, comparing and contrasting, defining, enumerating, or through a combination of methods.
- Exposition explains why or how.
- Exposition promotes reader understanding of a subject.

ELEMENT 1

General Competence

Read each essay as a whole, first, in order to form an overall judgement of its quality. To assign the essay a score, consider the following questions: To what extent does the essay achieve an expository purpose for the intended audience? To what extent does the essay organize its elements to create an effective whole?

Assign each paper a mark of 1-6 using the scale below:

MASTER

6 = Very competent	The paper executes all the elements competently. There are no serious errors. The paper has a clear main idea, logical organization, relevant, detailed support, and a command of basic mechanics. There are no major flaws.
5 = Definitely competent	The paper is competent in all of the basic elements, but there may be a few minor flaws.
4 = Adequately competent	The paper is adequately competent in all elements. There may be a few flaws. Some may be serious.*

NON-MASTER

3 = Almost competent	The paper lacks competence in one or two elements, and there are several major flaws.
2 = Not very competent	The paper lacks competence in two or more of the elements. There are many serious flaws.
1 = Not at all competent	Paper has none or only one of the elements competently executed.

---

* If the essay is only one paragraph, paragraph cohesion is not considered a missing element, if the one existing paragraph coheres. If it clearly should have been divided into several paragraphs, then paragraph cohesion is a missing element.

ELEMENT 2

Essay Organization and Coherence

This subscale focuses on the flow of ideas throughout the entire paper and between paragraphs. The emphasis is on vertical relationships of ideas throughout the essay.

The paper has a main idea (stated or clearly implied) which makes a point about the subject and is at a greater level of generality than the other points within the paper. Subtopics are logically related to the main idea and to each other.

MASTER

6 = ° The subject is identified.
° The main idea is stated or implied in opening and/or closing statement.
° Opening and closing statements match or logically relate to the text and to each other.
° The topic is limited through reference to key points or lines of reasoning.
° The essay plan is logical.
° The essay plan is consistently maintained (no digression or extraneous material).

5 = ° The subject is identified.
° The main idea is stated or implied in opening and/or closing statement.
° Opening and closing statements relate to or follow from the text and from each other.
° The topic is partly limited by indicating number and type of key points.
° Plan is logical.
° Plan is signalled by appropriate transitions.
° There may be digression or an elaboration.

4 = ° The subject is identified.
° The main idea is identified or implied.
° There may or may not be an attempt to limit the topic, give directions to subsequent reasoning.
° There may be a few minor digressions from the plan, but no major digressions.
° Subtopics can be reshuffled.

NON-MASTER

3 = ° Subject is clear.
° Main point may not be very clear. There may be a major digression or several minor digressions.
° A plan is attempted which may need to be inferred.

2 = ° Subject is clear.
° Main idea is not very clear and/or there may be more than one.
° The plan is attempted, but not consistently or not completely carried out.
° There are many digressions.

1 = ° Subject is unclear.
° Main idea is absent or very unclear.
° No plan is attempted or followed.

ELEMENT 3

Organization - Paragraph*

This subscale focuses on the relationship of ideas within paragraphs; their logical interrelationship and subordination to the paragraph topic.

Paragraphs present subtopics which are developed by cohesive groups of supporting statements. Each subtopic represents a complete unit of thought. Major units of thought are delineated by physical separation of paragraphs. Statements within the paragraph relate logically to each other and to the paragraph subtopic.

MASTER

6 = ° All major units of thought are set off by distinct paragraphs.
° The paragraph has a clearly stated or implied topic.
° All sentences within paragraphs are related to each other and to the paragraph topic, and are subordinate to it. There are no digressions or irrelevancies. There are no one-sentence paragraphs unless they are especially effective.

5 = ° Most major subtopics are developed in paragraphs.
° Most paragraphs contain logically related subordinate support.
° There may be a minor digression.

4 = ° Many subtopics are developed in discrete paragraphs with related subordinate support.
° There may be some minor digressions.

NON-MASTER

3 = ° In some paragraphs statements are logically related, but may not function as support subordinate to the paragraph topic. Paragraph separation is evident and consistent. Some relationships between sentences must be inferred.

2 = ° There are few paragraphs where statements are logically related or supported. There are many digressions. Paragraph separation is inconsistent. Many relationships among sentences must be inferred.

1 = ° There are no paragraphs where statements logically cohere. Paragraph separation is incorrect.

---

* For conventions for paragraph separation (e.g., physical separation or indenting) see Mechanics.

ELEMENT 4

Support

This subscale focuses on the quality (specificity and relationship) of the support provided vertically for the essay theme as well as horizontally within each paragraph.

Support statements are at a greater level of specificity than the generalizations they are intended to support. Support statements logically relate to each other and to the generalization. Support includes specific details such as examples, facts, anecdotes, reasons, and concrete language.

MASTER

6 = ° The essay's main idea and all paragraph topics are supported by relevant, specific statements.

5 = ° The essay's main idea and almost all paragraph generalizations/assertions are supported by predominantly specific statements. Enumerations are supported by descriptive detail, functions, or rationale.

4 = ° The essay's main idea and most paragraph generalizations are supported. Most support is specific. Enumerations are supported by descriptive detail, functions, or reasons.

NON-MASTER

3 = ° Some or all generalizations are supported by logically related detail, or some support is not specific but it is distinct and clear. Support may be primarily an unelaborated, undetailed, unsupported list.

2 = ° An attempt is made to support generalizations/assertions. Some supporting sentences do not logically follow from each other or are redundant.
° Support lacks precision, clarity in details, and/or language.

1 = ° No support is provided, or,
° Support, if present, is vague, and confusing, or,
° Not logically related to generalizations, or,
° At the same level of generality as the topic it attempts to support.

ELEMENT 5

Grammar/Mechanics (Usage, sentence construction, spelling, punctuation, capitalization)

Errors in grammar or mechanics are considered according to how seriously they interfere with communication. These errors may be global or local. A global error makes it difficult to understand the writer's message. The sentence, "Tomorrow, I went to the store," for example, forces the reader to decipher which time context (future or present) the writer is actually alluding to. A local error does not seriously interfere with the writer's message. For example, in the sentence, "He going to the store now," the message is clear but the grammar is incorrect. Naturally, an overabundance of errors which, if individually considered are local, can seriously distract the reader's attention and understanding. The intent here is to evaluate errors in relation to how much they interfere with the writer's effectiveness in communicating rather than to attempt to assign different values to the myriad of possible grammatical and mechanical errors that can occur.

MASTER

6 = ° There are few or no errors. There are no serious errors.

5 = ° There may be a few minor errors in the categories, but no more than one serious error.

4 = ° There are some errors. A few may be serious.

NON-MASTER

3 = ° There are numerous errors in the categories. There are some serious errors in several categories. Below mastery in sentence construction.

2 = ° There are many serious errors, causing some confusion.

1 = ° Errors are so numerous and serious that they interfere with communication.

Check those mechanical skills below master level.

_____ Usage. Does not display command of standard vocabulary usage.

_____ Sentence construction. Does not display command of basic sentence structure.

_____ Spelling. Misspells many common words (includes homonyms).

_____ Capitalizations and punctuation. Does not use standard conventions appropriately, e.g., periods, commas, capitals, apostrophes.

_____ Conventions of paragraph separation are incorrect or inconsistent (indenting, spacing, titles, numbers).

1) Usage

Serious errors:

- Homonyms, e.g., it, it's; their, there; to, two, too
- Incorrect use of common words
- Incorrect pronoun reference

Minor errors:

- awkward or odd use of words, phrases, but meaning still clear
- vague, abstract language

2) Sentence construction

Serious Errors:

- Subject verb agreement
- Run on
- Fragments

3) Spelling

Serious errors:

- Common words misspelled; does not include homonyms. Any misspelled word only counts as one error, even if the misspelling repeats.

Minor errors:

- Unusual, less frequent words

4) Capitalization/Punctuation

- Initial capitals -- common proper nouns
- Periods at end of sentence, common abbreviations
- Commas (in series, for opening phrases)
- Contractions

5) Paragraph Conventions

Serious errors:

- Title
- Number
- Inconsistency of separation convention
- Absence of any convention for separation

Minor errors:

- Use of spacing instead of indentation (as in business letters)

Narrative Scale

Global Rating Procedures

The purpose of global rating is to form an overall judgement of how well a piece of writing communicates a message to the reader. Global scoring assumes that each element of an essay is integrally related to other elements in the essay and the effect is cumulative. In short, the procedure views a piece of writing as a total work, the whole being greater than the sum of its parts.

For the element of "General Competence," you are being asked to form an overall judgement concerning a composition's effectiveness as an example of narrative writing. All of the elements presented in the scale--focus/organization, support, and grammar/mechanics--should be considered in rating for general competence as well as any other elements which you may feel are important.

ELEMENT 1

General Impression

Read each essay as a whole, first, to form an overall judgement of its quality. To assign the essay a score, consider the following questions: To what extent is the essay an example of effective narration? To what extent does the essay organize its elements to create an effective whole?

MASTER

6 = An excellent example of narration. Each element is evident in use throughout the essay (e.g., topic is clearly identified, characters defined, situations fully developed).

5 = A good example of narration, but the elements are not equally well developed throughout the essay/paragraph.

4 = An adequate example of narration. The writer incorporates the elements of narration. The essay is simple, informing and clear, and presents nothing more than essentials. There may be one or two isolated instances of global error and no more than three local errors or a total of five local errors.

NON-MASTER

3 = A marginal example. The writer presents evidence of limited skill in using the narrative elements. The elements are all present but developed poorly. Extremes are noted: strong to weak subject focus, proper to poor grammar and mechanics.

2 = A poor example. Writer's use of elements is problematic, focus on topic steadily decreases, support statements, if present, are weak; grammatical and mechanical errors also present.

1 = The paper is either off topic or genre. Presents no central subject, supports are irrelevant, or absent, and contains numerous grammatical and mechanical errors. The essay cannot be comprehended.

ELEMENT 2

Focus/Organization

This subscale examines whether the topic is clearly indicated and developed in an organized manner. The composition should exhibit a clear structure both within and between paragraphs. The topic should be limited and free of extraneous material.

MASTER

6 = ° The subject of the composition is clearly stated or implied.
° The topic is clearly limited; there are no digressions or extraneous material.
° Events and/or ideas are presented in a logical manner; they are linked both within and between paragraphs.
° The subject is developed through description of events, setting, or through the thoughts, emotions of the characters involved.

5 = ° The subject of the composition is clearly stated or implied.
° There may be one or two brief digressions or elaborations, but the topic is clearly limited.
° Events and/or ideas are presented in a logical manner; they are smoothly linked both within and between paragraphs.
° The subject is developed through description of events, setting, or through the thoughts and/or emotions of the characters involved.

4 = ° The subject of the composition is clearly stated or implied.
° There may be one or two brief digressions or elaborations, but the topic is clearly limited.
° Events and/or ideas are presented in a logical manner, but linking within and/or between paragraphs may be weak in one or two instances.
° The subject is developed through description of events, setting, or through the thoughts and/or emotions of the characters involved.
° The writer gives evidence of knowing the concepts of focus and organization.

NON-MASTER

3 = ° The subject of the composition is stated or implied.
° The logic or progression of ideas/events within or between paragraphs is often unclear; linking is frequently weak.
° The topic development through events, setting, or character thoughts/feelings is uneven; there may be too much or too little elaboration of some aspects.

2 = ° The subject of the composition is stated or implied, but the main point is not clear.
° There are many digressions or elaborations; the topic is not limited.
° The logic or progression of ideas and/or events is so unclear that no story line is discernible.
° There is little development through events, setting, or characters' feelings or thoughts.

1 = ° The subject of the composition is unclear or absent.
° There is no obvious organizational plan.
° There is no development of events, setting, or characters.
° There is much extraneous material.

ELEMENT 3

Support

This subscale focuses on the quality (specificity and relationship) of the support provided for the essay theme both within each paragraph as well as throughout the essay as a whole.

Support statements should be at a greater level of specificity and depth than the generalizations they are intended to develop. Events, descriptions, and characters should be developed through the use of specific, well-integrated details such as examples, facts, anecdotes, or descriptions. These details should provide the reader with an image of the appearance, feelings, thoughts, actions, or mood of the events taking place in the narrative.

MASTER

6 = ° Events, characters, and/or descriptions are developed by specific and clear supporting details, such as examples, descriptions, anecdotes, facts, etc.
° Supporting details provide an image/feeling of actions, appearance, feelings, thoughts, and/or mood in the paragraph/essay.
° Supporting details are more specific than the general ideas/events; characters are described and well integrated to the rest of the paragraph/essay.

5 = ° Events, characters, and/or descriptions are developed by specific and clear supporting details, such as examples, descriptions, anecdotes, facts, etc.
° The use of supporting detail is not consistent throughout. For example, in one or two instances, the writer may not provide information about appearance, feelings, thoughts, actions, or mood.
° Although most supporting statements or details provide in-depth descriptions and are more specific than the general statements they describe, there may be one or two instances where the detail lacks depth.

4 = ° Except for one or two instances, events, characters, and actions are described through the use of adjectives, adverbs, prepositional phrases, etc.
° Events, characters, and/or actions are developed by supporting detail.
° The use of supporting detail may be inconsistent or rudimentary. One aspect mentioned in the essay (i.e., character, event, or description) may not be sufficiently developed. Overall, however, the writer gives evidence of using supporting detail to develop most aspects of the essay/paragraph.

NON-MASTER

3 = ° The use of detail is very uneven. Several statements/descriptions are not developed through the use of detail.
° There is too much superficial or irrelevant detail.
° Some of the detail is not well integrated within the paragraph/essay.

2 = ° There is too little detail. There are very few instances where supporting detail is used to develop events, characters, or descriptions.
° Most details lack specificity and depth.
° The details are not smoothly integrated in the composition.

1 = ° Supporting details are vague, confusing, or not related to the events, characters, or descriptions they are meant to describe.
° There is little or no evidence of supporting details in the paragraph/essay.

ELEMENT 4

Grammar/Mechanics

Errors in grammar or mechanics are considered according to how seriously they interfere with communication. These errors may be global or local. A global error makes it difficult to understand the writer's message. The sentence, "Tomorrow, I went to the store," for example, forces the reader to decipher which time context (future or present) the writer is actually alluding to. A local error does not seriously interfere with the writer's message. For example, in the sentence, "He going to the store now," the message is clear but the grammar is incorrect. Naturally, an overabundance of errors which, if individually considered are local, can seriously distract the reader's attention and understanding. The intent here is to evaluate errors in relation to how much they interfere with the writer's effectiveness in communicating rather than to attempt to assign different values for the myriad of possible grammatical and mechanical errors that can occur.

MASTER

6 = ° The writer appears to have control of the usage and mechanical aspects of this essay.
° There are no global errors. Recall that global errors affect the essay/paragraph as a whole and interfere with clarity of communication.
° There are only one or two local errors (if any) in mechanics. For example, there may be a few spelling errors of difficult words, e.g. antenna.

5 = ° Usage and mechanics are not a problem in this paper.
° There are only a few local errors in usage or mechanics.
° There may be one isolated global error, but the general meaning is clear throughout the paragraph/essay.

° Usage and mechanics are not a problem in this paper.
° Errors do not interfere with the clarity of communication. For example, confusion of to, too, two; their, there; or other local errors.
° There may be one or two isolated instances of global errors, and no more than three local errors, or a total of five local errors.

NON-MASTER

3 = ° Some errors, global and local, do interfere with the clarity of communication. For example, there may be a long run-on sentence, inappropriate fragments, or incorrect tense continuity (e.g., Yesterday we are going to school.")

2 = ° Errors may detract from the clarity of communication, such as confusing antecedents, omission of key words, serious misspellings of common words (e.g., confurmable/comfortable, laike/like).
° There are many global and local errors throughout the essay/paragraph.

1 = ° Errors make this paper very difficult to read and understand. There is an overabundance of global errors and a significant amount of communication is lost.
° Errors are not restricted to one type of problem, such as run-on sentences.

Mechanics/Grammar
Examples of Errors

I. Sentence Construction

Global errors

- tense continuity (e.g., Yesterday I go to ...)
- very long run-on sentences
- fragments

Local errors

- subject/verb agreement (He go to the store.)
- short run-ons
- incorrect or lack of connectors between clauses

II. Usage

Global errors

- incorrect use of common words
- code switching (mixing languages) (Today I played with my sister y fuimos to the store.)

Local errors

- homonyms (e.g., it, it's; their, there; to, two, too)
- incorrect pronoun reference (That is she book.)
- awkward or odd use of words, phrases, but meaning still clear

III. Spelling

Global errors

- Common words frequently misspelled; does not include homonyms. Any misspelled word only counts as one error, even if the misspelling repeats (comfortable-confurbal; should not-shurent; wrong-rong)
- Words misspelled so seriously as to impede communication (e.g., mild down/melted down)

Local errors

- unusual, less frequent words

IV. Punctuation/Capitalization

Global errors

- periods at the end of sentences
- contractions
- commas where understanding is impeded

Local errors

- initial caps -- common proper nouns
- commas (in series, for opening phrases)
- periods for abbreviation

TABULAR EXPOSITORY SCALE

General Competence	Focus/Organization	Support	Grammar/Mechanics
6 Excellent (all)	• subject clear • main idea clear • key points, reasoning at beginning or end • beginning and end relate • plan logical • plan signaled by transitions • paragraphs set off all major ideas • no one sentence paragraphs • no digressions	• main idea and all major topics elaborated by specific details • enumerations supported by detail, functions, rationale	• one or two minor errors • no gross errors
5 Good (most)	• subject clear • main idea clear • topic partly limited by reference to number or type of key reasons • logical plan • some transitions • most major ideas in paragraphs • minor digression	• almost all major points elaborated by specific details • most elaboration is specific	• a few minor errors • may be one gross error • usage and mechanics still not a problem
4 Adequate (many)	• subject clear • main idea clear • topic may be limited • plan logical, but sub-topic can be reshuffled • many major thoughts in paragraphs • few minor digressions, no major digressions	• many major points supported • much elaboration is specific	• a few common errors • one or two gross plus no more than one minor error
3 Developing (some)	• subject clear • main idea not very clear • plan attempted, must infer • some logically developed paragraphs • some major digressions or excessive elaboration	• some major points developed • some elaboration is specific but is distinct and clear • may be a list	• some errors interfere • some gross and minor errors • sentence construction below mastery
2 Rudimentary (few)	• subject clear • main idea not very clear or more than one • plan attempted • few paragraphs logically developed • many digressions	• supportive detail attempted • may be redundant • may not be precise or clear	• many gross and minor errors • some confusion
1 Off topic Off genre (almost none)	• subject may be unclear • main idea unclear • plan unclear • almost no logically developed paragraphs	• little or no support or • support is confusing or at the same level of generality as the main assertion	• difficult to read • many gross, varied errors • very confusing

APPENDIX B

Student Compositions Rated During a Training Session

Samples of Student Compositions Rated During a CSE Training Session

The five student writing samples in this appendix were rated during a CSE rater training session for secondary school teachers (Quellmalz, 1982). The samples are representative 10th-grade student exposition, and were among those used early in the training as a check on how uniformly and consistently raters were applying the criteria.

The rated writing samples show the level of agreement between two rater trainees. They also show how closely each seemed to be following the criteria agreed upon, rather than applying perhaps more idiosyncratic values.

At the top of each student composition are two numbers. These numbers show the General Competence rating that the two raters independently assigned to the sample. These general impression ratings indicate each rater's assessment of how well the student's writing met the specific skills required in Focus/Organization, Support, and Grammar/Mechanics.

Each sample composition is followed by a feedback sheet showing a CSE rater's judgment of its general competency level, as well as of how well it dealt with the skills addressed in the three remaining subscales.

As is readily seen, there are high levels of agreement between the two raters, and similarly acceptable agreement between the raters and the assessments provided on the feedback sheets. Had these agreement levels not been achieved, the feedback sheets would have been used as a focus for discussing the causes of rater disagreement and for resolving differences before taking up the next set of compositions to be rated.

10th Grade Prac. | 5 | 4 | 10-2

Promises are made to be kept. I feel that if someone makes a promise and then they break it, it is disloyal or in a way, lying. If a person is unsure that he or she can keep a certain promise, they shouldn't make it!

I have this one special friend who asked a girl to the Homecomming dance at our school. She agreed to go along with him. About one week after he asked her, she told him she couldn't go because she got grounded. He was furious. One of the girls friends told me that she just didn't want to go with him so she made up that lie. If this is not a good example of breaking a promise, I don't know what is. She had promised to go along with him and then just dumped him, so to speak.

This little story shows that breaking a promise is not very nice, it is actually very bad. Alot of the time a person will have to make up a story to break a promise. Breaking promises is an example of bad taste for the breaker, and it usually hurts the breakie.

Set Number Training Check 1 (2)

Code Number _____

FEEDBACK SHEET

ELEMENT	RATING	
General Competency	4	The main point is clear and specifically supported. Some mechanical errors and awkward usage.
Focus/Organization	4	Paper has main idea summarized in the opening and closing paragraphs. No use of transitions.
Support	4	Although there is only one example, it is clear and specific.
Grammar/Mechanics	4	There are a few minor errors. "If a person... they." "Brakie."

Mechanical Skills below mastery

_____ Sentence Construction
_____ Usage
_____ Spelling
_____ Punctuation and Capitalization
_____ Paragraph Conventions

1/82

| 4 | 4 |

10

10th Grade Prompt

What is a promise? A promise is something one says, and assumed to be true. An example is, if you promise to clean the table off after diner your parents assume your statement to be true, and will expect your promise to be followed through. ~~with~~ A promise is your own word, and when you make a promise, it is important to keep it. ~~If you~~ Promises should never be broken. If you promise something, you should never go back on your word. If you don't think you can't keep a promise you are about to make, why even make it? By breaking promises, people won't trust you. One example of why promises should not be broken, is as follows: Sally was a very popular high school student, ~~these~~ she was getting ready for a date one day, when her friend Patty was over. As she was going to get her make-up from her purse, a pack of pills fell out. when Patty asked Sally what they were, Sally told Patty, she would tell her, if she promised not to say anything to anyone

to one. Patty promised, and Sally preceeded to tell her they were birth control pills, that she was taking, because Bill, her boyfriend, was putting pressure on her. The next day at school, Patty was talking with her friends. While talking, she accidently told about Sally and her birth control pills. When Sally finds out, do you think she will trust Patty anymore? ~~Probably~~ Surely not, because now she doesn't trust her, and can't confide her secrets in her.

Promises shouldn't be said, if you don't feel you can keep them. People who break promises, will not be trusted or looked upon with respect. If you want to be respected and trusted, think before you promise.

Set Number Training Check 6 (7)

Code Number _____

FEEDBACK SHEET

ELEMENT	RATING	
General Competency	3	Paper is generally well developed, but flawed by poor paragraphing and numerous mechanical errors.
Focus/Organization	3	Position is clear and logically developed, but paragraphing is a problem. Good conclusion.
Support	5	Gives a specific example to support thesis.
Grammar/Mechanics	3	Numerous awkward usages. Problem with use of commas.

Mechanical Skills below mastery

_____ Sentence Construction
✓ Usage
_____ Spelling
✓ Punctuation and Capitalization
_____ Paragraph Conventions

Rev. 1/82

We use promises during our regular day almost all the time. Some promises we are conscious of and some we are not conscious of. We often promise people that we will do things so they will have more confidence in us. Like when we borrow money from someone, you almost always say "I promise I will pay you back!" promises are also broken almost as fast as they are made. Very often promises should be broken though. Sometimes you can get caught in the middle of two people and might have to. To break a promise because a close friend may be close to getting hurt

Set Number Training Check 7 (9)

Code Number _____

FEEDBACK SHEET

ELEMENT	RATING	
General Competency	2	The line of reasoning is confusing. Support is vague and mechanics are intrusive.
Focus/Organization	2	The main idea is unclear and there is no conclusion. Paper moves from "definition of promises" to "very often they should be broken." No evidence of paragraphing.
Support	2	Support is attempted, but not specific.
Grammar/Mechanics	3	Sentence errors. Some awkward usage. "We use promises..." "Like when..." "Get caught in the middle of two people."

Mechanical Skills below mastery

- ✓ Sentence Construction
- _____ Usage
- _____ Spelling
- _____ Punctuation and Capitalization
- _____ Paragraph Conventions

Rev. 1/82

#5

FOR US A PROMISE IS ONE OF THE MOST IMPORTANT THINGS. TO KEEP I THOUGHT ABOUT KEEPING YOUR PROMISES ETC. IF YOU DON'T FULFILL YOUR PROMISES TO ONE WHO BELIEVE YOU IS THEY THINK YOUR ALWAYS LIEING.

ONE DAY THERE WERE TWO GUYS WHO ALWAYS TELL US TO SOMEBODY WE AFTERNOON IT WAS VERY HOT. THEY ESCAPE TO TAKE A SHOT WHAT SAW IS TURN ON THE AIR CONDITION. AFTER FEW MINUTE THE AIR CONDITION STARTED BURNING ALL THEY CAN'T GO OUT THE DOOR BECAUSE IT WAS BURN STARTING TO BURN. THE ONLY WAY TO GET OUT IS TO CALL THE FIRE DEPARTMENT ACROSS THE STREET. THEY GRAB THE TELEPHONE & TRIED TO CALL THE FIRE DEPARTMENT BUT THEY DIDN'T BECAUSE BECAUSE THOSE WORKS WERE THEY THOUGHT THEY WERE LIELS.

THEREFORE IN ORDER PEOPLE WILL BES YET YOUR WORD KEEP YOUR PROMISES.

Set Number Training Check 4 (5)

Code Number _____

FEEDBACK SHEET

ELEMENT	RATING
General Competency	2 Main point confusing because of poor logic and numerous mechanical errors.
Focus/Organization	2 The logic is unclear. The paper confuses promises with lying.
Support	2 Support is attempted, but unclear.
Grammar/Mechanics	1 Major usage, sentence construction and spelling errors interfere with meaning.

Mechanical Skills below mastery

- ✓ Sentence Construction
- ✓ Usage
- ✓ Spelling
- _____ Punctuation and Capitalization
- _____ Paragraph Conventions

Rev. 1/82

4 | 3 |

10 3

PROMISES ARE BIG THINGS AND SMALL THINGS, IT DEPENDS ON HOW YOU LOOK AT IT. I TRY NOT TO MAKE PROMISES BECAUSE THE PERSONS ASKING ME TO MAKE A PROMISE USUALLY USES IT AS A BLACK MAIL

FOR EXAMPLE THIS GIRL I USED TO GO OUT WITH IS CONSTANTLY PESTERING ME TO MAKE PROMISES LIKE SHE WANTED ME TO TEACH HER HOW TO SURF AND TAKE HER DOWN TO THE BEACH JUST SO SHE COULD TRY AND SINK INTO MY HEAD ONCE AGAIN.

THERE ARE TIMES LIKE THAT WHEN I LIKE TO BREAK PROMISES JUST FOR MY OWN GOOD OTHERWISE I WOULD GET SCREWED UNDER BY HER ONCE AGAIN. FOR NOW ON I'M GOING TO TRY AND AVOID MAKING PROMISES TO ANYONE

GO

Set Number Training Check 2 (3)

Code Number _____

FEEDBACK SHEET

ELEMENT	RATING
General Competency	3 The papers' point is at times confused by imprecise language.
Focus/Organization	2 Logical development unclear. "promise = blackmail" "promise = are big and small things"
Support	2 Support is not very specific or clear.
Grammar/Mechanics	3 Usage is a pervasive problem. "For now on" "Like she wanted me to. . ." slang terms. Subject-verb agreement.

Mechanical Skills below mastery

_____ Sentence Construction
✓ Usage
_____ Spelling
_____ Punctuation and Capitalization
_____ Paragraph Conventions

Rev. 1/82